

1. Какая проблема, связанная с несбалансированными выборками, может возникнуть в подходе "один-против-всех" для многоклассовой классификации? **1 / 1 point**

- ☐ Данный подход не будет работать на несбалансированных выборках.
- ☐ На несбалансированных выборках данный подход будет работать слишком долго.
- ☒ Каждая задача бинарной классификации может оказаться несбалансированной, если классов много.

✓ **Correct**

Если в выборке 1000 классов одинакового размера, то в каждой бинарной задаче размер положительного класса будет составлять 0.1% от всей выборки — а это очень мало.

2. В каком случае может понадобиться переход в спрямляющее признаковое пространство? **1 / 1 point**

- ☐ Наличие коррелирующих признаков.
- ☐ Наличие шумовых признаков.
- ☒ Нелинейная зависимость ответа от признаков.

✓ **Correct**

Возможно, есть спрямляющее пространство, в котором линейные методы будут работать хорошо.

3. В лекциях обсуждалось, что несбалансированные выборки являются проблемой из-за того, что методы обучения классификаторов тем или иным образом минимизируют долю неправильных ответов. Эта метрика никак не различает разные виды ошибок, и поэтому с её точки зрения константный классификатор может оказаться лучшим выбором. Почему же тогда не решить проблему использованием чувствительной метрики — например, точности выделения объектов маленького класса? **1 / 1 point**

- ☒ Точность — дискретная метрика, её невозможно оптимизировать градиентными методами.

✓ **Correct**

- ☐ Точность (precision) и подобные ей метрики не позволяют обнаружить плохое качество выделения маленького класса.
- ☒ Oversampling и undersampling, по сути, приводят к оптимизации точности путём выставления больших весов ошибкам определённого типа.

✓ **Correct**